Numer. Math. 68: 403–425 (1994)

Numerische Mathematik
© Springer-Verlag 1994

Pseudozeros of polynomials and pseudospectra of companion matrices*

Kim-Chuan Toh[1], **Lloyd N. Trefethen**[2]

[1] Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA
e-mail: kc@cam.cornell.edu
[2] Department of Computer Science, Cornell University, Ithaca, NY 14853, USA
e-mail: LNT@cs.cornell.edu

Received June 15, 1993

Summary. It is well known that the zeros of a polynomial p are equal to the eigenvalues of the associated companion matrix A. In this paper we take a geometric view of the conditioning of these two problems and of the stability of algorithms for polynomial zerofinding. The ϵ-*pseudozero set* $Z_\epsilon(p)$ is the set of zeros of all polynomials $\hat{p}$ obtained by coefficientwise perturbations of p of size $\le \epsilon$; this is a subset of the complex plane considered earlier by Mosier, and is bounded by a certain generalized lemniscate. The ϵ-*pseudospectrum* $\Lambda_\epsilon(A)$ is another subset of $\mathbb{C}$ defined as the set of eigenvalues of matrices $\hat{A} = A + E$ with $\|E\| \le \epsilon$; it is bounded by a level curve of the resolvent of A. We find that if A is first balanced in the usual EISPACK sense, then $Z_{\epsilon\|p\|}(p)$ and $\Lambda_{\epsilon\|A\|}(A)$ are usually quite close to one another. It follows that the Matlab ROOTS algorithm of balancing the companion matrix, then computing its eigenvalues, is a stable algorithm for polynomial zerofinding. Experimental comparisons with the Jenkins-Traub (IMSL) and Madsen-Reid (Harwell) Fortran codes confirm that these three algorithms have roughly similar stability properties.

Mathematics Subject Classification (1991): 65F35

1. Introduction

Zeros of polynomials and eigenvalues of nonsymmetric matrices are well-known examples of problems whose answers may be highly sensitive to perturbations. The sensitivity of these two problems was made famous among numerical analysts by Wilkinson in the early 1960s, and contributed to his development of the notions of stability and conditioning that are now standard in this field [19, 22, 23]. And, of course, the two problems are related, for the zeros of a polynomial are the same as the eigenvalues of the associated companion matrix.

Despite the classical nature of the subject, the relationship between these two problems has received less study than one might suppose. Polynomial zerofinding has been something of a backwater in numerical analysis, and it is probably fair to

* Supported by NSF Grant DMS-9116110 and DOE Grant DE-FG02-94ER25199

say that although all numerical analysts know that one can find zeros via companion matrices in principle, most assume that it isn't a good idea to do so. It seems to be not well known, though it is pointed out in the original paper on the subject [8], that the central step of the Jenkins-Traub algorithm for polynomial zerofinding is equivalent to a generalized Rayleigh quotient iteration for finding an eigenvalue of the companion matrix (for details see the Appendix).

The approach we take in this paper is a geometric one. For a monic polynomial $p(z)$, let $Z(p)$ denote the zero set (= set of zeros) of $p(z)$, and, for any $\epsilon \geq 0$, define the *ϵ-pseudozero set* of $p(z)$ by

(1) $$Z_\epsilon(p) = \{z \in \mathbb{C} : z \in Z(\hat{p}) \text{ for some } \hat{p}\},$$

where $\hat{p}$ ranges over polynomials whose coefficients are those of p modified by perturbations of size $\leq \epsilon$ (for details see Section 2). The relevance of such sets to the conditioning of the zerofinding problem has been discussed by Mosier [16]. Similarly, for a matrix A, let $\Lambda(A)$ denote the spectrum of A, and define the *ϵ-pseudospectrum* of A by

(2) $$\Lambda_\epsilon(A) = \{z \in \mathbb{C} : z \in \Lambda(A + E) \text{ for some } E \text{ with } \|E\| \leq \epsilon\}.$$

Matrix pseudospectra have been studied by Trefethen [21], Godunov [4], and others going back at least to H.J. Landau in 1975 [11].

In this paper we report numerical experiments that show that $Z_{\epsilon\|p\|}(p)$ and $\Lambda_{\epsilon\|A\|}(A)$ are generally quite close to one another when A is a companion matrix of p that has been "balanced" in the usual EISPACK sense proposed originally by Parlett and Reinsch [18]. It follows that the polynomial zerofinding problem and the balanced matrix eigenvalue problem are comparable in conditioning and therefore that finding zeros via eigenvalues of companion matrices, the method used by the Matlab ROOTS command, is a stable algorithm. To test these conclusions we compare ROOTS with the Jenkins-Traub (IMSL) code CPOLY [10] and the Madsen-Reid (Harwell) code PA16 [13] for a variety of polynomials. Our experiments suggest that all three codes are reliable, with the highest accuracy typically achieved by PA16 and the next-best accuracy by ROOTS.

The significance of pseudozero sets and pseudospectra is not just a matter of rounding errors and numerical stability. In any applied mathematical problem that apparently depends on polynomial zeros, it is likely that what really matters is whether $|\,p(z)\,|$ is very small, not necessarily exactly zero. Similarly, in a matrix eigenvalue problem, what really matters may be whether $\|(zI - A)^{-1}\|$ is very large, not necessarily exactly infinity. Thus the investigation of pseudozero sets and pseudospectra is a natural extension of the zerofinding and eigenvalue problems themselves, not just a tool for analyzing numerical algorithms.

2. Pseudozero sets of polynomials

We begin this section by introducing some notation:

$\mathbb{P}$: The set of monic polynomials of degree n with complex coefficients. For each $p \in \mathbb{P}$, we will express p as

$$p(z) = \sum_{i=0}^{n} c_i z^i, \quad c_n = 1.$$

We will also denote the vector of coefficients $(c_0, \ldots, c_{n-1})^{\mathrm{T}}$ by p when there is no danger of confusion.

$Z(p)$: The set of zeros of p.

p_*: The reciprocal polynomial of p, i.e., $p_*(z) = z^n p(z^{-1})$.

$\mathscr{D}$: The set of $n \times n$ diagonal matrices with diagonal entries in $\mathbb{C}$. For each $D \in \mathscr{D}$, we denote its diagonal vector $(d_0, \ldots, d_{n-1})^T$ by d; d^{-1} denotes the vector $(d_0^{-1}, \ldots, d_{n-1}^{-1})^{\mathrm{T}}$.

$\|x\|_d$: For each nonsingular $D \in \mathscr{D}$, the norm $\|x\|_d$ is defined over $\mathbb{C}^n$ by

$$\|x\|_d = \|Dx\|_2, \quad x \in \mathbb{C}^n.$$

$\tilde{z}$: For each $z \in \mathbb{C}$, $\tilde{z}$ denotes the vector $(1, z, \ldots, z^{n-1})^{\mathrm{T}}$.

$e_1, \ldots, e_n$: the standard basis in $\mathbb{C}^n$.

For a fixed $D \in \mathscr{D}$, we can assign a metric on $\mathbb{P}$ by

$$\|p - \hat{p}\|_d = \left[\sum_{i=0}^{n-1} |\, d_i\,|^2 |\, c_i - \hat{c}_i\,|^2 \right]^{1/2},$$

which measures the perturbations in the coefficients of p relative to the weights given by d. Note also that the norm $\|x\|_d$ induces a matrix norm on $\mathbb{C}^{n \times n}$ (the space of all $n \times n$ matrices over $\mathbb{C}$), which satisfies the formula $\|A\|_d = \|DAD^{-1}\|_2$.

For each $p \in \mathbb{P}$, we define the *ϵ-pseudozero set* of $p(z)$ by

(3) $$Z_\epsilon(p; d) = \{z \in \mathbb{C} : z \in Z(\hat{p}) \text{ for some } \hat{p} \in \mathbb{P} \text{ with } \|\hat{p} - p\|_d \leq \epsilon\}.$$

These sets quantify the conditioning of the zerofinding problem and arise naturally in analyzing the stability of zerofinding algorithms. For a zerofinding algorithm to be stable, the computed zeros of p should lie in a region $Z_{Cu}(p; d)$ for an appropriate choice of d, where $C = O(\|p\|_d)$ and u is the machine precision. For example, the choice $d = \|p\|_2 p^{-1}$ corresponds to coefficientwise perturbations and $d = \sqrt{n}(1, \ldots, 1)^{\mathrm{T}}$ corresponds to normwise perturbations.

The following proposition gives an algebraic characterization of $Z_\epsilon(p; d)$ in terms of the level curves of a certain function. This allows us to determine $Z_\epsilon(p; d)$ numerically.

Proposition 2.1.

$$Z_\epsilon(p; d) = \left\{ z \in \mathbb{C} : \frac{|\, p(z)\,|}{\|\tilde{z}\|_{d^{-1}}} \leq \epsilon \right\}.$$

Mosier [16] introduced these sets and proved this proposition in the ∞-norm. Here we extend the result to the 2-norm; in fact, it is valid in the p-norm for any $1 \leq p \leq \infty$.

Proof. If $z \in Z_\epsilon(p; d)$, then $\exists\ \hat{p} \in \mathbb{P}$ with $\hat{p}(z) = 0$ and $\|\hat{p} - p\|_d \leq \epsilon$. By the Hölder inequality,

$$\begin{aligned} |\, p(z)\,| &= |\, \hat{p}(z) - p(z)\,| = |\sum (\hat{c}_i - c_i) z^i\,| \\ &\leq \|\hat{p} - p\|_d\ \|\tilde{z}\|_{d^{-1}} \leq\ \epsilon \|\tilde{z}\|_{d^{-1}}. \end{aligned}$$

To show the converse, consider $z \in \mathbb{C}$ such that $|\, p(z)\,| \leq \epsilon \|\tilde{z}\|_{d^{-1}}$. Let $\theta = \arg(z)$. Consider the degree $n-1$ polynomial r defined by

$$r(w) = \sum_{k=0}^{n-1} r_k w^k, \text{ where } r_k = |z^k| \, \mathrm{e}^{-\mathrm{i}k\theta} d_k^{-2}.$$

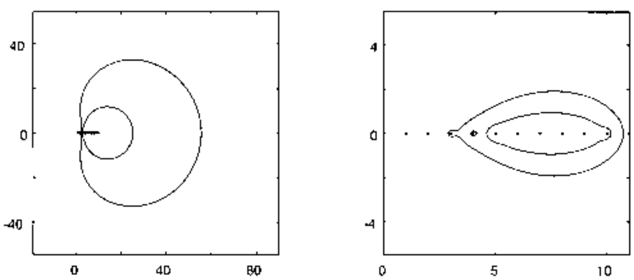
Then $r(z) = \|r\|_d \|\tilde{z}\|_{d^{-1}}$, and the polynomial $\hat{p} \in \mathbb{P}$ defined by $\hat{p}(w) = p(w) - \frac{p(z)}{r(z)}\, r(w)$ satisfies $\hat{p}(z) = 0$ and $\|\hat{p} - p\|_d = \frac{|p(z)|}{|r(z)|}\|r\|_d \le \epsilon$. Thus $z \in Z_\epsilon(p; d)$. □

Note that if $d = e_1{}^{-1} = (1, \infty, \ldots, \infty)^{\mathrm{T}}$, then $Z_\epsilon(p; d)$ becomes the region bounded by the usual *ϵ-lemniscate* of p [6],

$$L_\epsilon(p) = \{z \in \mathbb{C} : |p(z)| \le \epsilon\}.$$

More generally we may call the boundary of $Z_\epsilon(p; d)$ a *generalized lemniscate*.

Example 2.1. Consider the monic polynomial p of degree 10 with zeros $1, 2, \ldots, 10$. Figure 1 shows the $\epsilon\|p\|_d$-pseudozero sets of p for two different choices of d corresponding to coefficientwise ($d = \|p\|_2 p^{-1}$) and normwise ($d = \sqrt{n}(1, \ldots, 1)^{\mathrm{T}}$) perturbations.



(a) normwise (b) coefficientwise

Fig. 1. $\epsilon\|p\|_d$-pseudozero sets ($\epsilon = 10^{-7}, 10^{-6}$) corresponding to normwise and coefficientwise perturbations for the degree-10 monic polynomial with zeros $1, 2, \ldots, 10$. Note the different scales in (a) and (b)

The ideas described above apply to arbitrary polynomials and finite perturbations. For the limiting case of infinitesimal perturbations, a single condition number suffices to describe what may happen to each simple root. Let the condition number of the root ξ of p be defined by

$$\kappa(\xi, p; d) = \lim_{\|\hat{p}-p\|_d \to 0} \sup_{\hat{p}} \frac{|\hat{\xi} - \xi|}{\|\hat{p} - p\|_d / \|p\|_d}.$$

Then from Proposition 2.1 we can readily derive the following formula:

Proposition 2.2.

$$\kappa(\xi, p; d) = \|p\|_d \frac{\|\tilde{\xi}\|_{d^{-1}}}{|p'(\xi)|}. \tag{4}$$

(Thus for infinitesimal ϵ, the component of $Z_{\epsilon\|p\|_d}(p; d)$ about ξ is the disk of radius $\epsilon\kappa(\xi, p; d)$). This result is essentially equivalent to formulas obtained by Gautschi in a sequence of several papers on polynomial condition numbers [3]. We define the condition number of the zerofinding problem for p to be

$$\kappa(p; d) = \max_{\xi} \; \kappa(\xi, p; d)\;. \tag{5}$$

3. Pseudospectra of companion matrices

For each monic polynomial $p = \sum_{i=0}^{n} c_i z^i \in \mathbb{P}$, the companion matrix associated with p is the $n \times n$ matrix

$$A_p = \begin{pmatrix} 0 & \cdots & \cdots & -c_0 \\ 1 & \ddots & & -c_1 \\ & \ddots & \ddots & \vdots \\ & & 1 & -c_{n-1} \end{pmatrix}. \tag{6}$$

The characteristic polynomial of A_p is p itself, and therefore the set of eigenvalues of A_p coincides with the set of zeros of p. We now introduce concepts for companion matrices analogous to those just given for polynomials.

For each $p \in \mathbb{P}$, we define the *ϵ-pseudospectrum* of A_p by

(7) $\Lambda_\epsilon(A_p; d) = \{z \in \mathbb{C} : z \in \Lambda(\hat{A}) \text{ for some } \hat{A} \in M_n \text{ with } \|\hat{A} - A_p\|_d \le \epsilon\}$.

The size of $\Lambda_\epsilon(A_p; d)$ is related to the conditioning of the companion matrix eigenvalue problem. The appearance of the norm $\|\cdot\|_d$ in (7) corresponds to the consideration of diagonally weighted perturbations in the entries of A_p:

$$\|\hat{A} - A_p\|_d = \|D(\hat{A} - A_p)D^{-1}\|_2. \tag{8}$$

The main reason we include D in our formulation is that the process of balancing a matrix, to be discussed in the next section, involves a diagonal similarity transformation; thus this formulation allows us to treat the balanced and unbalanced cases together. For an eigenvalue algorithm applied to the companion matrix to be stable, the computed eigenvalues of A_p should lie in a region $\Lambda_{Cu}(A_p; d)$ for some $C = O(\|A_p\|_d)$.

The following proposition gives an algebraic characterization of $\Lambda_\epsilon(A_p; d)$ in terms of the level curves of the norm of the resolvent of A_p.

Proposition 3.1.

$$\Lambda_\epsilon(A; d) = \{z \in \mathbb{C} : \|(zI - A)^{-1}\|_d \ge \epsilon^{-1}\}.$$

This result holds for any $n \times n$ matrix and any matrix norm induced by a vector norm; see [1]. For completeness, we include a proof, omitting the subscript d for simplicity.

Proof. It is easily shown that $\Lambda_\epsilon(A_p; d)$ is a subset of the right-hand side. To show the converse, consider $z \in \mathbb{C}$ such that $\|(zI - A)^{-1}\| = {\epsilon_0}^{-1}$ for some $\epsilon_0 \le \epsilon$. We can find $u \in \mathbb{C}^n$ such that

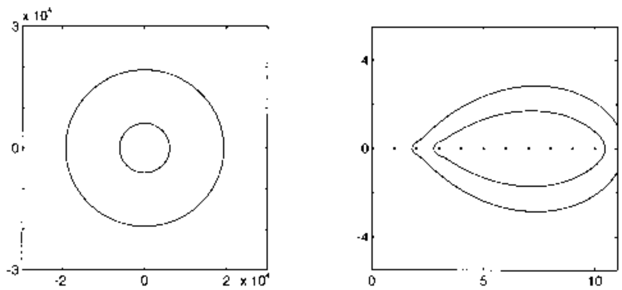
$$\|(zI - A)^{-1}u\| = {\epsilon_0}^{-1} \text{ and } \|u\| = 1.$$

Let $v = \epsilon_0(zI - A)^{-1}u$; then $\|v\| = 1$. By a standard corollary of the Hahn-Banach theorem, there exists a linear functional $f \in (\mathbb{C}^n)^*$ (the dual space of $\mathbb{C}^n$) such that $f(v) = 1$ and $\|f\| = 1$. Let $a = (f(e_1), \dots, f(e_n))^*$; then $f(x) = a^*x \quad \forall x \in \mathbb{C}^n$ (the superscript $*$ means conjugate transpose). Consider the matrix $\hat{A} = A + \epsilon_0 u a^*$. $\hat{A}$ satisfies

$$\hat{A}v = Av + \epsilon_0 u a^* v = Av + \epsilon_0 u = zv,$$

showing that z is an eigenvalue of $\hat{A}$. Now $\|\hat{A} - A\| = \|\epsilon_0 u a^*\| = \epsilon_0\|f\| = \epsilon_0 \le \epsilon$, so $z \in \Lambda_\epsilon(A_p; d)$, and this completes the proof. $\square$

Example 3.1. Consider the companion matrices corresponding to the same polynomial p as in Example 2.1. Figure 2 compares $\epsilon\|A_p\|_d$-pseudospectra of the unbalanced and balanced (see next section) companion matrices, which turn out to be far larger in the unbalanced case. By contrast, note the approximate agreement of the curves in Fig. 2b with those in Fig. 1b.



(a) unbalanced (b) balanced

Fig. 2. $\epsilon\|A_p\|_d$-pseudospectra ($\epsilon = 10^{-7}, 10^{-6}$) of the unbalanced and balanced companion matrices corresponding to the same polynomial as in Fig. 1. Note the different scales in (a) and (b), and the approximate agreement of Figs. 1b and 2b

We define the condition number of the simple eigenvalue λ of a matrix B by

$$\kappa(\lambda, B; d) = \lim_{\|\hat{B}-B\|_d \to 0} \sup_{\hat{B}} \frac{|\hat{\lambda} - \lambda|}{\|\hat{B} - B\|_d / \|B\|_d}.$$

A formula for this quantity is (see [5]):

$$\kappa(\lambda, B; d) = \|B\|_d \frac{\|x\|_{d^{-1}} \|y\|_d}{|x^* y|}, \tag{9}$$

where the superscript $*$ denotes conjugate transpose and x and y are left and right eigenvectors of B, respectively, corresponding to the eigenvalue λ. (Just as (4) is a consequence of Proposition 2.1, (9) can be derived as a consequence of Proposition 3.1.) For the special case of the companion matrix A_p, the left and right eigenvectors can be written in closed form:

$$x = (1, \lambda, \ldots, \lambda^{n-1})^{\mathrm{T}}, \tag{10}$$

$$y = (b_0, b_1, \ldots, b_{n-1})^{\mathrm{T}}, \tag{11}$$

where $b_0, \ldots, b_{n-1}$ (dependent on λ) are the coefficients of the monic polynomial $(p(z) - p(\lambda))/(z - \lambda) = \sum_{i=0}^{n-1} b_i z^i$. The expression for the condition number then reduces to the following formula:

Proposition 3.2.

$$\kappa(\lambda, A_p; d) = \|A_p\|_d \frac{\|\tilde{b}(\lambda)\|_d \|\tilde{\lambda}\|_{d^{-1}}}{|p'(\lambda)|}, \tag{12}$$

where $\tilde{b}(\lambda) = (b_0, \ldots, b_{n-1})^{\mathrm{T}}$.

(Thus for infinitesimal ϵ, the component of $\Lambda_{\epsilon\|A_p\|_d}(A_p;d)$ about λ is the disk of radius $\epsilon\kappa(\lambda, A_p; d)$). If the eigenvalues of A_p are simple (equivalently, the zeros of p are simple), we define the condition number of the problem of finding the eigenvalues of A_p with respect to perturbations of the form given in (8) to be

$$\kappa(A_p; d) = \max_{\xi}\ \kappa(\xi, A_p; d). \tag{13}$$

4. The importance of balancing

The conditioning of the eigenvalue problem for a companion matrix A_p may be changed enormously by a diagonal similarity transformation. The best one could hope for would be to find such a transformation that makes the eigenvalue problem no worse conditioned than the underlying polynomial zerofinding problem. Our experiments show that to up to a factor of 10 or so, the similarity transformation known as *balancing* tends to achieve just this. In this section we shall give some details, comparing balancing with other diagonal similarity transformations and measuring the results by the scalar condition number κ of (13). The plots of the next section carry these observations further, revealing that balancing tends to achieve an approximate match not only of condition numbers but also of pseudospectra and pseudozero sets.

Recall from (8) that problems of the condition of A_p can be formulated in two equivalent ways: we can either leave the 2-norm fixed and change A_p to DA_pD^{-1}, where D is a diagonal matrix, or leave A_p fixed and change the 2-norm to $\|\cdot\|_d$, where $d = \text{diag}(D)$. Using the latter formulation, we consider four possibilities:

1. Pure companion matrix. One possibility is to consider the matrix (8) with no diagonal similarity transformation. Equivalently, the diagonal similarity transformation is defined by $d = e = \sqrt{n}(1, \ldots, 1)^{\mathrm{T}}$.

2. Balancing. Balancing the matrix A_p corresponds to finding a diagonal matrix $T \in \mathscr{D}$ such that TA_pT^{-1} has the 2-norm of its ith row and ith column approximately equal for each $i = 1, \ldots, n$ (we denote $\|\text{diag}(T^{-1})\|_2\ \text{diag}(T)$ by t). This idea was introduced in [18] and is a standard option in EISPACK [20] and also the default procedure in Matlab [14].

3. Scaling. For each scalar $\alpha > 0$, the associated scaled monic polynomial of $p \in \mathbb{P}$ is

$$p_\alpha(z) = \frac{1}{\alpha^n}\ p(\alpha z) = \sum_{i=0}^{n} \frac{c_i}{\alpha^{n-i}} z^i.$$

The diagonal similarity transformation that corresponds to this scaling operation is defined by the diagonal matrix $D_\alpha = \text{diag}(d^{(\alpha)})$, where

$$d^{(\alpha)} = \|(\alpha^n, \ldots, \alpha^1)\|_2(\alpha^{-n}, \ldots, \alpha^{-1})\ .$$

4. Coefficientwise. If p has all nonzero coefficients, there is a natural diagonal similarity transformation associated with it, viz., the coefficientwise diagonal similarity transformation, which is represented by the diagonal matrix $C = \text{diag}(c)$, where $c = \|p\|_2 p^{-1}$.

We have found empirically that the balancing operation tends to achieve the best conditioned eigenvalue problem for A_p among these four choices of diagonal similarity transformations. Specifically, consider the ratios of the eigenvalue condition numbers for the three other problems to that of the balanced companion matrix, i.e., the ratios

$$\sigma(A_p; d) = \frac{\kappa(A_p; d)}{\kappa(A_p; t)} \tag{14}$$

for $d = c$, $d^{(\alpha)}$, and e, with t corresponding to the balancing operation as indicated above. In the case of polynomial scaling, α is chosen to be the optimal value in the sense that $\sigma(A_p; d^{(\alpha)})$ is minimized. We have found that $\sigma(A_p; e)$ and $\sigma(A_p; d^{(\alpha)})$ are usually much greater than unity, implying that these two eigenvalue problems are much worse conditioned than the eigenvalue problem associated with the balanced matrix. The third ratio, $\sigma(A_p; c)$, is of order 1 in many cases, but coefficientwise transformations have the defect that they are not well defined when some of the coefficients of p are zero.

Having found that balancing operation achieves approximately the best conditioned eigenvalue problem for A_p, we then compare this matrix condition number with the condition number of the coefficientwise perturbed zerofinding problem for p. That is, we consider the ratio

$$\frac{\kappa(A_p; t)}{\kappa(p; c)}.$$

Our experiments show that this ratio tends to be of order 1.

Table 1 gives empirical evidence to support these statements. It is a tabulation of the four ratios discussed above for a variety of degree-20 monic polynomials:

(1) "Wilkinson polynomial": zeros $1, 2, 3, \ldots, 20$.
(2) the monic polynomial with zeros equally spaced in the interval $[-2.1, 1.9]$.
(3) $p(z) = \sum_{k=0}^{20} z^k/k!$.
(4) the Bernoulli polynomial of degree 20.
(5) the monic polynomial $z^{20} + z^{19} + z^{18} + \cdots + 1$ (zeros are the 21st roots of unity except 1).
(6) the monic polynomial with zeros $2^{-10}, 2^{-9}, 2^{-8}, \ldots, 2^{9}$.
(7) the Chebyshev polynomial of degree 20.
(8) the monic polynomial with zeros equally spaced on a sine curve, viz., $(2\pi/19(k + 0.5)) + \mathrm{i}\sin(2\pi/19(k + 0.5))$, $k = -10, -9, -8, \ldots, 9$.

The scatter plot of $\kappa(p; c)$ versus $\kappa(A_p; t)$ in Figure 3 provides further evidence that the condition of the balanced companion matrix eigenvalue problem is typically comparable to that of the coefficientwise perturbed zerofinding problem. For this plot, we considered a random sample of one hundred degree-10 monic polynomials with coefficients of the form

$$a_1 \times 10^{e_1} + \mathrm{i}a_2 \times 10^{e_2}, \tag{15}$$

where a_i and e_i ($i = 1, 2$) are drawn from the uniform distributions on the intervals $[-1, 1]$ and $[-10, 10]$, respectively. The idea of considering polynomials with random coefficients of this form to test the quality of zerofinding algorithms was proposed by Jenkins and Traub in 1974 [9].

Table 1. Comparison of the conditioning of the balanced companion matrix eigenvalue problem with that of the eigenvalue problems of various matrices diagonally similar to A_p

	coefficientwise	scaled	(α)	unbalanced	$\kappa(A_p;t)/\kappa(p;c)$
(1)	3.4×10^0	1.9×10^8	(9.0)	2.3×10^{31}	1.4×10^1
(2)	2.0×10^0	4.2×10^5	(0.9)	4.5×10^5	3.6×10^0
(3)	1.4×10^0	1.7×10^2	(10.0)	1.1×10^{32}	5.2×10^0
(4)	$*$	4.6×10^3	(1.6)	2.9×10^6	8.7×10^0
(5)	1.1×10^0	9.4×10^{-1}	(1.1)	1.1×10^0	7.6×10^0
(6)	1.7×10^1	1.2×10^{24}	(0.8)	1.4×10^{24}	2.1×10^1
(7)	$*$	6.4×10^3	(0.6)	6.5×10^4	4.4×10^0
(8)	$*$	3.4×10^4	(1.7)	9.6×10^7	2.6×10^0

*: coefficientwise transformation not defined because of zero coefficient

We should note that the conditions of the balanced companion matrix eigenvalue problem and the coefficientwise perturbed zerofinding problem are closer for some polynomials than others. When the zeros of the polynomials are of widely varying magnitudes, the match tends to be worse; this accounts for the deviation of some points from the dashed line in Fig. 3.

5. Numerical experiments

We now come to a sequence of more detailed experiments, Figs. 4–11. For the same eight polynomials described in the last section, each figure presents first of all a graphical comparison of two pseudozero sets $Z_{\epsilon\|p\|_c}(p;c)$ ($c = \|p\|_2 p^{-1}$) of the polynomial with the corresponding pseudospectra $\Lambda_{\epsilon\|A_p\|_t}(A_p;t)$ of the associated balanced companion matrix. A reasonably close agreement is observed in all cases.

This agreement of pseudozero sets and pseudospectra suggests that it ought to be possible to compute zeros of polynomials stably via eigenvalues of balanced companion matrices. To test this prediction, we have compared three zerofinding methods:

1. J-T. Our first zerofinder is the Jenkins-Traub program CPOLY, available from ACM TOMS via Netlib and also in the IMSL library [8]. The innermost step of this algorithm is equivalent to a certain modified generalized Rayleigh quotient iteration for companion matrices (see the Appendix). A three stage procedure is used, each stage being characterized by the type of shift involved.

2. M-R. Our second zerofinder is the Madsen-Reid code PA16 from the Harwell Library [13], which is a Newton-based method coupled with line search. The algorithm is designed to find zeros beginning with the zero of smallest magnitude of a given p. Stage 1 of the method uses Newton formula to find a search direction for the real function $|p(z)|$. Once the iterates are close enough to a zero of $p(z)$, the algorithm enters stage 2 and uses the standard Newton iteration.

3. ROOTS. Finally, we consider ROOTS, the Matlab zerofinding code based on constructing the balanced companion matrix and finding its eigenvalues by standard methods [15]. (Matlab actually uses the QZ algorithm, but our experiments suggest that the QR algorithm gives similar results.)

In our experiments, we first find the "exact" roots of p by computing the eigenvalues of A_p in quadruple precision ($\approx$ 34 digits) via a standard EISPACK computational

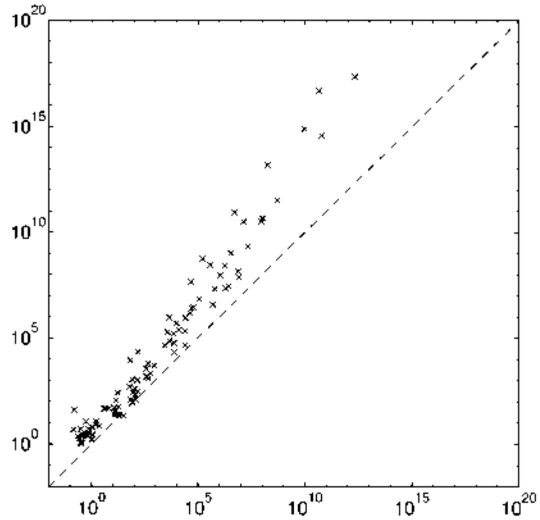


Fig. 3. Comparison of the condition number $\kappa(A_p; t)$ of the balanced companion matrix eigenvalue problem with the condition number $\kappa(p; c)$ of the coefficientwise perturbed zerofinding problem for 100 degree-10 polynomials with random coefficients of the form (15)

path for eigenvalues of a general complex matrix, (CBAL, CORTH, COMQR) [20]. The rest of our computations are then carried out in double precision ($\approx$ 16 digits).

For each of the eight polynomials, Figs. 4–11 list the maximum absolute error of the roots computed by the above algorithms together with the condition numbers of the coefficientwise perturbed zerofinding problem for p and the balanced companion matrix eigenvalue problem for A_p.

Figures 4–11, and our other numerical experiments, indicate that PA16 is typically the most accurate of these zerofinding algorithms, followed by ROOTS. Further empirical evidence is presented in Figs. 12–14, which show scatter plots corresponding to the same sample of one hundred random degree-10 monic polynomials as in Fig. 3. A few remarks are in order concerning these scatter plots:

First, the scatter plots indicate that ROOTS and PA16 are stable. The Jenkins–Traub code CPOLY appears somewhat unstable in a few cases.

Second, the error in the zeros computed by ROOTS is typically of the order of magnitude of the condition number of the coefficientwise perturbed zerofinding problem, not the balanced companion matrix eigenvalue problem. With reference to Figure 3, we thus observe that ROOTS finds the zeros somewhat more accurately than would be predicted by the condition number of the balanced companion matrix eigenvalue problem. We have carried out numerical experiments to explain this phenomenon and find that it is apparently a consequence of the sparsity of the balanced companion matrices. The condition number (9) is of course defined for a general square matrix, not taking sparsity or other structure into account.

Finally, the reader will note the appearance of some points well below the dashed diagonal line for J-T and M-R, but not ROOTS. These correspond to examples with zeros that are nearly multiple. Evidently the performance of ROOTS matches linear

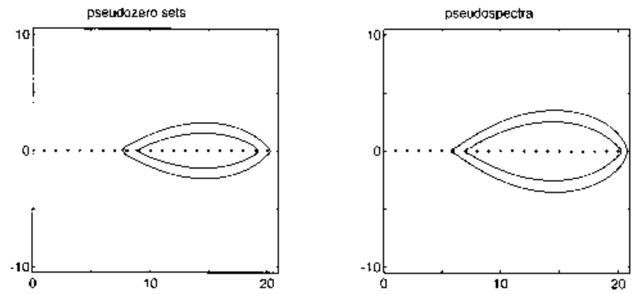


Fig. 4. Wilkinson polynomial (ϵ = $10^{-14}, 10^{-13}$)

error(J-T)	=	$1.36e-02$	$u\ \kappa(p;c)$	=	$2.76e-01$
error(M-R)	=	$1.53e-02$	$u\ \kappa(A_p;t)$	=	$4.00e+00$
error(ROOTS)	=	$3.58e-03$			

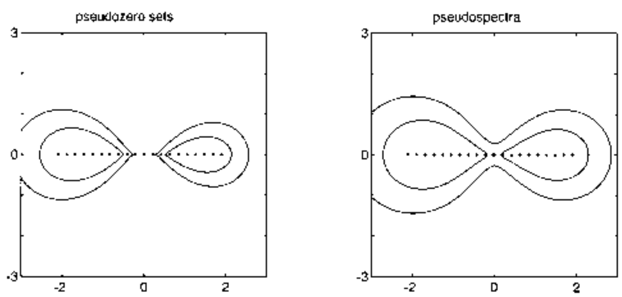


Fig. 5. Polynomial with zeros equally spaced in $[-2.1, 1.9]$ (ϵ = $10^{-3}, 10^{-2}$)

error(J-T)	=	$5.31e-13$	$u\ \kappa(p;c)$	=	$6.57e-12$
error(M-R)	=	$2.46e-13$	$u\ \kappa(A_p;t)$	=	$2.34e-11$
error(ROOTS)	=	$6.11e-13$			

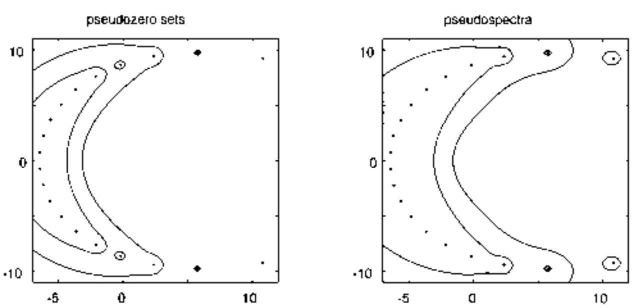


Fig. 6. Partial sums of the Taylor series for e^z (ϵ = $10^{-4}, 10^{-3}$)

error(J-T)	=	$1.98e-11$	$u\ \kappa(p;c)$	=	$3.16e-11$
error(M-R)	=	$1.00e-12$	$u\ \kappa(A_p;t)$	=	$1.66e-10$
error(ROOTS)	=	$9.00e-12$			

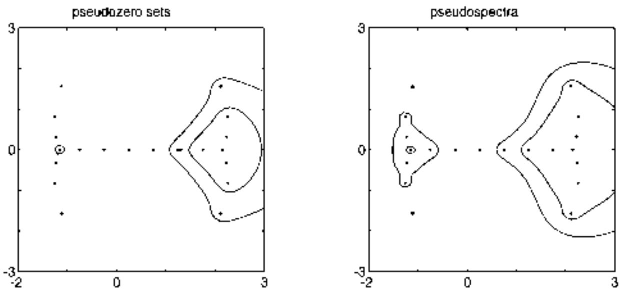


Fig. 7. Bernoulli polynomial ($\epsilon = 10^{-4}, 10^{-3}$)

error(J-T)	=	$4.10e-12$	$u\ \kappa(p;c)$	=	$1.10e-11$
error(M-R)	=	$1.17e-12$	$u\ \kappa(A_p;t)$	=	$9.62e-11$
error(ROOTS)	=	$1.38e-12$			

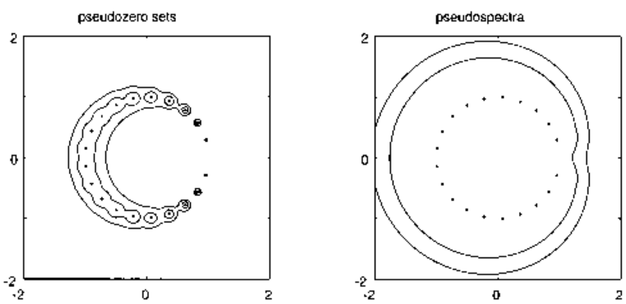


Fig. 8. Polynomial with zeros equal to the 21st roots of unity except 1 ($\epsilon = 10^{-1.2}, 10^{-1}$)

error(J-T)	=	$4.92e-13$	$u\ \kappa(p;c)$	=	$4.22e-16$
error(M-R)	=	$2.48e-16$	$u\ \kappa(A_p;t)$	=	$3.19e-15$
error(ROOTS)	=	$1.29e-15$			

perturbation theory in these cases, whereas J-T and M-R actually do somewhat better. Thus in these cases ROOTS could be said to be mildly unstable.

6. Asymptotic agreement of pseudozero sets and pseudospectra

In this section we establish various mathematical relationships between pseudozero sets and pseudospectra of the associated companion matrices, showing in particular that in the limits $\epsilon \to 0$ (infinitesimal perturbations) and $\alpha \to \infty$ (zeros at the origin), they become identical to one another.

Let us begin by noting that the ϵ-pseudozero set of p, $Z_\epsilon(p;d)$, is contained in the ϵ'-pseudospectrum of A_p, $\Lambda_{\epsilon'}(A_p;d)$, if $\epsilon' = \epsilon\,|d_{n-1}|^{-1}$. This is true because we may view $Z_\epsilon(p;d)$ as the set of eigenvalues obtained by structured perturbations of A_p that preserve the companion matrix form:

$$Z_\epsilon(p;d) = \{z \in \mathbb{C} \;\; z \in \Lambda(A_{\hat{p}}) \text{ for some } \hat{p} \in \mathbb{P} \text{ with } \|A_{\hat{p}} - A_p\|_d \le \epsilon'\}.$$

For a similar reason, the lemniscatic region $L_{\epsilon|d_0|^{-1}}(p)$ is contained in $Z_\epsilon(p;d)$, for we may view this region as the set of zeros obtained by perturbing only the constant

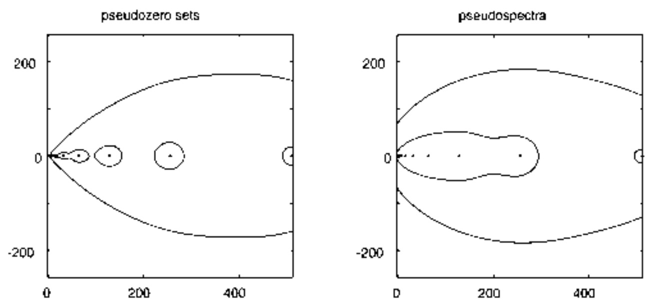


Fig. 9. Polynomial with zeros $2^{-10}, 2^{-9}, \ldots, 2^{9}$ ($\epsilon = 10^{-3}, 10^{-2}$)

error(J-T)	=	$5.12e-13$	$u\ \kappa(p;c)$	=	$6.49e-12$
error(M-R)	=	$2.84e-13$	$u\ \kappa(A_p;t)$	=	$1.36e-10$
error(ROOTS)	=	$9.66e-13$			

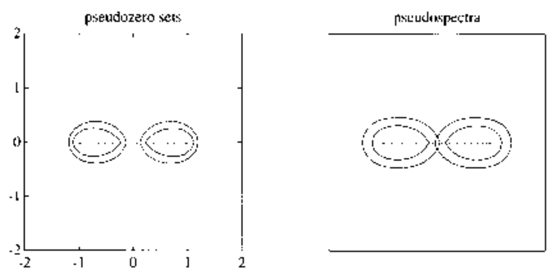


Fig. 10. Chebyshev polynomial $T_{20}(x)$ ($\epsilon = 10^{-4}, 10^{-3}$)

error(J-T)	=	$1.98e-10$	$u\ \kappa(p;c)$	=	$5.55e-11$
error(M-R)	=	$8.59e-12$	$u\ \kappa(A_p;t)$	=	$2.45e-10$
error(ROOTS)	=	$4.89e-12$			

term of p:

$$L_{\epsilon|d_0|^{-1}}(p) = \{z \in \mathbb{C}\ :\ z \in Z(p+\delta) \text{ for some scalar } \delta \in \mathbb{C} \text{ with } |\delta d_0| \le \epsilon\}.$$

In summary, we have the following proposition:

Proposition 6.1. *The lemniscatic regions, pseudozero sets and pseudospectra satisfy*

$$L_{\epsilon|d_0|^{-1}}(p) \subset Z_\epsilon(p;d) \subset \Lambda_{\epsilon|d_{n-1}|^{-1}}(A_p;d)$$

provided $d_0, d_{n-1} \neq 0$.

We showed in Sect. 2 that the ϵ-pseudozero set $Z_\epsilon(p;d)$ is the region bounded by the ϵ-level curve of a certain function defined over the complex plane (denoted by $\psi(\cdot\ ;d)$ later in this section). A similar result (Proposition 6.2) holds for the ϵ-pseudospectrum $\Lambda_\epsilon(A_p;d)$ when ϵ is sufficiently small (we denote the associated function, defined below, by $\phi(\cdot\ ;d)$). From these simple characterizations, we are able to give analytical relationships between:

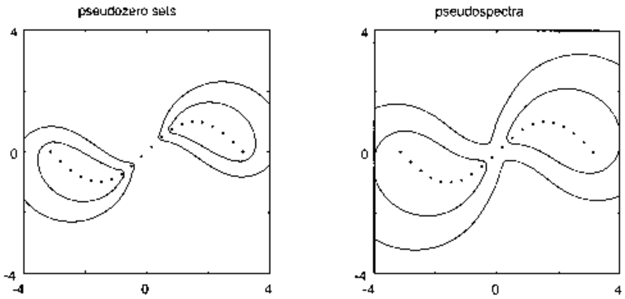


Fig. 11. Polynomial with zeros equally spaced on the section of the sine curve from $-\pi$ to π ($\epsilon = 10^{-3}, 10^{-2}$)

error(J-T)	=	$3.70e-13$	$u\ \kappa(p;c)$	=	$5.05e-12$
error(M-R)	=	$5.63e-13$	$u\ \kappa(A_p;t)$	=	$1.29e-11$
error(ROOTS)	=	$5.66e-13$			

1. *The pseudozero sets and the pseudospectra (limit $\epsilon \to 0$).* Corollary 6.1 shows that $\Lambda_\epsilon(A_p; d^{(1)})$ and $Z_{\epsilon'}(p; d^{(2)})$ agree with one another in the limit $\epsilon \to 0$, where

$$\epsilon' = \epsilon\ (\|p\|_{d^{(2)}}\kappa(A_p; d^{(1)}))/(\|A_p\|_{d^{(1)}}\kappa(p; d^{(2)})). \tag{16}$$

2. *The pseudospectra for different diagonal similarity transformations (limit $\epsilon \to 0$).* Corollary 6.2 shows that $\Lambda_\epsilon(A_p; d^{(1)})$ and $\Lambda_{\epsilon'}(A_p; d^{(2)})$ agree with one another in the limit $\epsilon \to 0$, where

$$\epsilon' = \epsilon\ (\|A_p\|_{d^{(2)}}\kappa(A_p; d^{(1)}))/(\|A_p\|_{d^{(1)}}\kappa(A_p; d^{(2)})). \tag{17}$$

3. *The lemniscatic regions, pseudozero sets and pseudospectra (limit $\alpha \to \infty$).* For the special case of polynomial scaling, Corollary 6.3 shows that the lemniscatic region $L_{\epsilon\alpha^{n-1}}(p)$, the pseudozero set $Z_{\epsilon\alpha^{n-1}}(p; d^{(\alpha)})$ and the pseudospectrum $\Lambda_\epsilon(A_p; d^{(\alpha)})$ agree with one another in the limit $\alpha \to \infty$.

Lemma 6.1. *For each $z \in \mathbb{C} - Z(p)$,*

$$(zI - A_p)^{-1} = \frac{1}{p(z)}\tilde{b}(z)\tilde{z}^{\mathrm{T}} - B(z) \tag{18}$$

$$= \frac{-1}{zp(z)}\tilde{h}(z^{-1})\tilde{z}^{\mathrm{T}} + H(z), \tag{19}$$

where $\tilde{b}(z) = (b_0(z), \ldots, b_{n-1}(z))^{\mathrm{T}}$ is the vector of coefficients of the polynomial $(p(w)-p(z))/(w-z) = \sum_{i=0}^{n-1} b_i(z)w^i$, $\tilde{h}(z^{-1}) = (h_{n-1}(z^{-1}), \ldots, h_0(z^{-1}))^{\mathrm{T}}$ is the vector of coefficients of the polynomial $(p_*(w) - p_*(z^{-1}))/(w - z^{-1}) = \sum_{i=0}^{n-1} h_i(z^{-1})w^i$, and

$$B(z) = \begin{pmatrix} 0 & 1 & \cdots & z^{n-2} \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}, \quad H(z) = z^{-1}\begin{pmatrix} 1 & & & \\ z^{-1} & \ddots & & \\ \vdots & \ddots & \ddots & \\ z^{1-n} & \cdots & z^{-1} & 1 \end{pmatrix}.$$

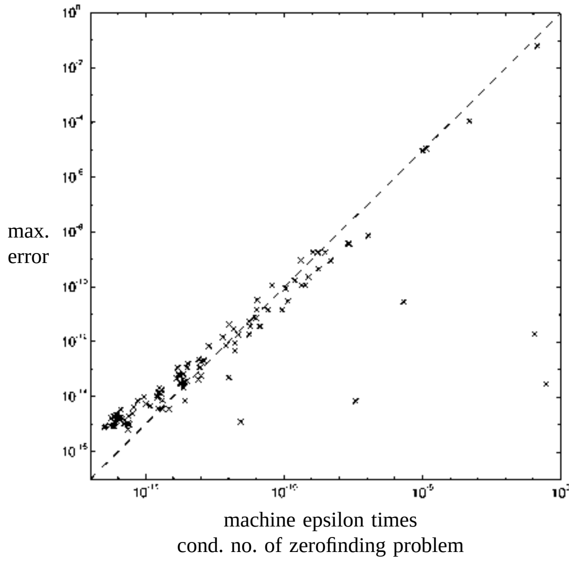


Fig. 12. Comparison of the accuracy of the Matlab companion matrix algorithm ROOTS with the conditioning of the coefficientwise perturbed zerofinding problem (100 random problems)

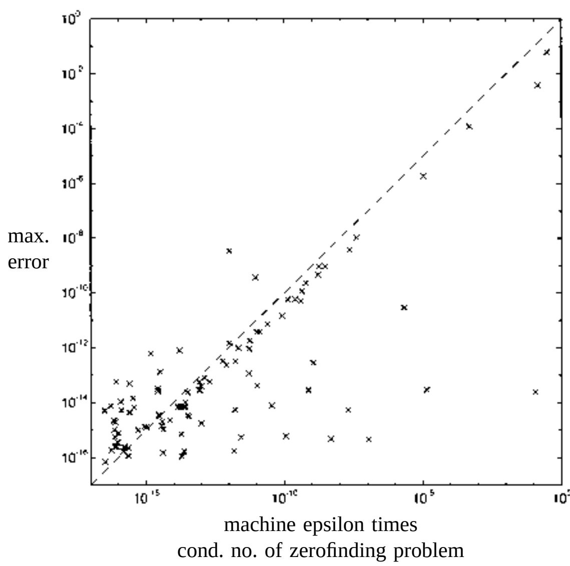


Fig. 13. Same as Fig. 12 but for the Jenkins-Traub code CPOLY

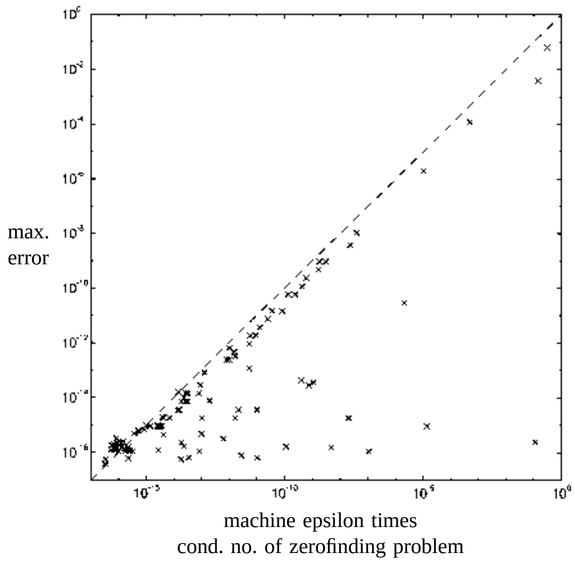


Fig. 14. Same as Fig. 12 but for the Madsen-Reid code PA16

Proof. We can write $zI - A_p$ as

$$zI - A_p = T(z) + pe_n^{\mathrm{T}},$$

where $T(z)$ is the bidiagonal matrix with z along its main diagonal and -1 along its subdiagonal. Application of the Sherman-Morrison formula gives (18), and further algebraic manipulations yield (19). □

For each $D \in \mathscr{D}$, let the functions $r(\cdot\,;d)$, $\psi(\cdot\,;d)$, and $\phi(\cdot\,;d)$ be defined by

$$r(z;d) = \begin{cases} \|\tilde{b}(z)\|_d & \text{if } |z| \le 1 \\ |z^{-1}|\, \|\tilde{h}(z^{-1})\|_d & \text{if } |z| > 1, \end{cases} \tag{20}$$

$$\psi(z;d) = \frac{|p(z)|}{\|\tilde{z}\|_{d^{-1}}}, \quad \phi(z;d) = \frac{r(z;d)}{\psi(z;d)}. \tag{21}$$

Note that $r(z;d)$ is continuous at each $z \in Z(p)$ because $\|\tilde{b}(z)\|_d = |z^{-1}|\, \|\tilde{h}(z^{-1})\|_d$ if $p(z) = 0$. Also, from Proposition 2.1, $Z_\epsilon(p;d)$ is the region bounded by the ϵ-level curve of $\psi(\cdot\,;d)$. Furthermore,

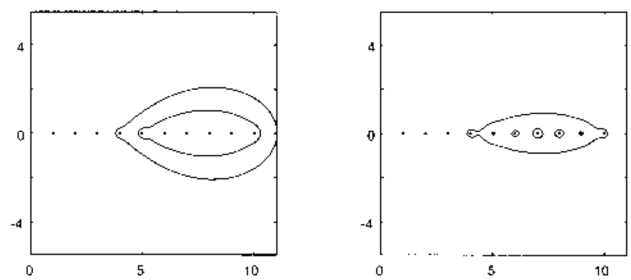
$$\kappa(z,p;d) = \|p\|_d \|\tilde{z}\|_{d^{-1}} / |p'(z)|, \tag{22}$$

$$\kappa(z,A_p;d) = \|A_p\|_d (r(z;d)\|\tilde{z}\|_{d^{-1}}) / |p'(z)|, \tag{23}$$

for all $z \in Z(p)$.

Proposition 6.2. *For each* $D \in \mathscr{D}$,

$$\frac{\|(zI - A_p)^{-1}\|_d}{\phi(z;d)} = 1 + O\left(\frac{1}{\phi(z;d)}\right) \quad \textit{as} \quad \phi(z;d) \to \infty. \tag{24}$$



normwise ($\epsilon = 10^{-10}$, 10^{-9}) coefficientwise ($\epsilon = 10^{-7}, 10^{-6}$)

Fig. 15. ϵ-pseudospectra of A_p (solid) and ϵ-level curves of $\phi(\cdot, d^{(i)})$ (dashed) corresponding to normwise and coefficientwise diagonal similarity transformations for the degree-10 monic polynomial with zeros $1, 2, \ldots, 10$. The dashed curves are not visible because they match the solid ones to plotting accuracy

Proof. From Lemma 6.1, for $|\, z \,| \le 1$,

$$\left| \|(zI - A_p)^{-1}\|_d - \frac{\|\tilde{b}(z)\|_d \|\tilde{z}\|_{d^{-1}}}{|\, p(z)\,|} \right| \le \|B(z)\|_d \le m_1,$$

where $m_1 = n \cdot \max\{\, |d_i/d_k| \;:\; 0 \le i \le k-1,\ 1 \le k \le n-1\}$. For $|\, z \,| > 1$, we have

$$\left| \|(zI - A_p)^{-1}\|_d - \frac{\|\tilde{h}(z^{-1})\|_d \|\tilde{z}\|_{d^{-1}}}{|\, zp(z)\,|} \right| \le \|H(z)\|_d \le m_2,$$

where $m_2 = n \cdot \max\{\, |d_i/d_k| \;:\; k+1 \le i \le n-1,\ 0 \le k \le n-2\}$. Combining these estimates gives

$$\left| \frac{\|(zI - A_p)^{-1}\|_d}{\phi(z; d)} - 1 \right| \le \frac{\max(m_1, m_2)}{\phi(z; d)},$$

which implies (24). □

Example 6.1. We may paraphrase Proposition 6.2 as follows: In the limit $\epsilon \to 0$, $\Lambda_\epsilon(A_p; d)$ agrees with the region bounded by the ϵ-level curve of $\phi(\cdot\;; d)$. Figure 15 illustrates this agreement pictorially. Consider the monic polynomial p with zeros $1, 2, \ldots, 10$ and diagonal similarity transformations $d^{(1)} = e$ and $d^{(2)} = c$. The figure shows the ϵ-level curves of $\|(zI - A_p)^{-1}\|_{d^{(i)}}$ and $\phi(z; d^{(i)})$, $i = 1, 2$. To the resolution of this figure, these two sets of curves are indistinguishable.

We note that in analyzing the conditioning of the eigenvalue problem of A_p, one is usually only interested in a region where $\phi(z; d) \gg 1$. In this case, Proposition 6.2 shows that $\phi(z; d)$ provides an accurate estimate of $\|(zI - A_p)^{-1}\|_d$. Note that only $O(n)$ flops are needed to calculate $\phi(z; d)$ by Horner's rule as compared with $O(n^3)$ flops needed to calculate $\|(zI - A_p)^{-1}\|_d$ by the SVD.

Corollary 6.1. *For any $D_1, D_2 \in \mathscr{D}$ and $z_0 \in Z(p)$,*

$$\frac{\|(zI - A_p)^{-1}\|_{d^{(1)}}}{1/\psi(z; d^{(2)})} \;\to\; \frac{\|p\|_{d^{(2)}}}{\|A_p\|_{d^{(1)}}} \frac{\kappa(z_0, A_p; d^{(1)})}{\kappa(z_0, p; d^{(2)})} \quad \textit{as} \quad z \to z_0. \tag{25}$$

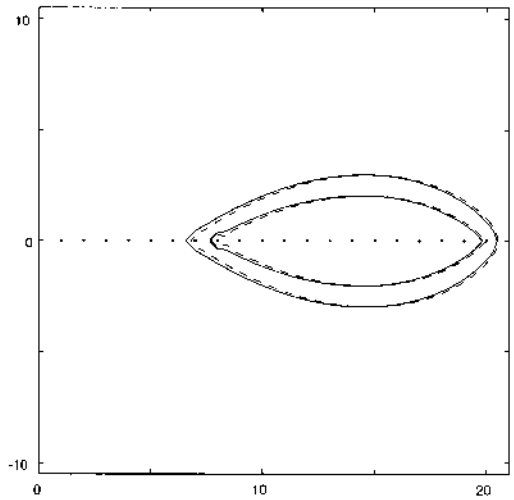


Fig. 16. ϵ-pseudospectra (solid) of the balanced companion matrix of A_p and ϵ'-pseudozero sets (dashed) of p with respect to coefficientwise perturbations for the monic polynomial with zeros $1, 2, \dots, 20$ ($\epsilon = 10^{-12}, 10^{-11}$)

Proof. The left-hand side is equal to

$$\frac{\|(zI - A_p)^{-1}\|_{d^{(1)}}}{\phi(z; d^{(1)})} \cdot \frac{r(z; d^{(1)})\|z\|_{1/d^{(1)}}}{\|z\|_{1/d^{(2)}}}.$$

Since $\phi(z; d) \to \infty$ as $z \to z_0$, by Proposition 6.2, the first fraction tends to 1 as $z \to z_0$. The second tends to

$$\frac{r(z_0; d^{(1)})\|z_0\|_{1/d^{(1)}}}{\|z_0\|_{1/d^{(2)}}}$$

because $r(z; d)$ is continuous at $z = z_0$. Substituting (22) and (23) into this limit gives (25). □

Example 6.2. We may paraphrase Corollary 6.1 as follows: In the limit $\epsilon \to 0$, the components of $\Lambda_\epsilon(A_p; d^{(1)})$ and $Z_{\epsilon'}(p; d^{(2)})$ (ϵ' as in (16)) containing z_0 agree with one another. Figure 16 illustrates this agreement for the "Wilkinson polynomial" p with zeros $1, 2, \dots, 20$. The figure compares the pseudospectra of the balanced matrix of A_p and the pseudozero sets of p with respect to coefficientwise perturbation. Choosing $z_0 = 15$, the figure plots the boundaries of $\Lambda_\epsilon(A_p; t)$ and $Z_{\epsilon'}(p; c)$ ($\epsilon' = \epsilon\,(\|p\|_c \kappa(z_0, A_p; t))/(\|A_p\|_t \kappa(z_0, p; c))$) for $\epsilon = 10^{-12},\ 10^{-11}$. The agreement is close.

Corollary 6.2. *For any $D_1, D_2 \in \mathscr{D}$ and $z_0 \in Z(p)$,*

$$\text{(26)} \qquad \frac{\|(zI - A_p)^{-1}\|_{d^{(1)}}}{\|(zI - A_p)^{-1}\|_{d^{(2)}}} \to \frac{\|A_p\|_{d^{(2)}}}{\|A_p\|_{d^{(1)}}} \frac{\kappa(z_0, A_p; d^{(1)})}{\kappa(z_0, A_p; d^{(2)})} \quad \textit{as} \quad z \to z_0.$$

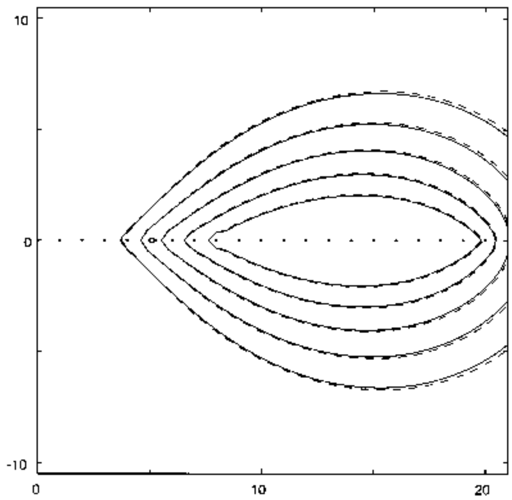


Fig. 17. ϵ-pseudospectra (solid) of the balanced matrix of A_p and ϵ'-pseudospectra (dashed) of the matrix similar to A_p via the coefficientwise diagonal similarity transformation for the same polynomial as in Fig. 16 ($\epsilon = 10^{-12}, \cdots, 10^{-8}$)

Example 6.3. Corollary 6.2 can be paraphrased as follows: In the limit $\epsilon \to 0$, the components of $\Lambda_\epsilon(A_p; d^{(1)})$ and $\Lambda_{\epsilon'}(A_p; d^{(2)})$ (ϵ' as in (17)) containing z_0 agree with one another. For illustration, we consider the "Wilkinson polynomial" again as in Example 6.2. For $z_0 = 15$, Fig. 17 plots the boundaries of $\Lambda_\epsilon(A_p; t)$ and $\Lambda_{\epsilon'}(A_p; c)$ for $\epsilon = 10^{-12}, \ldots, 10^{-8}$. Again the agreement is close.

Corollary 6.3. *For each scalar $\alpha > 0$, let $d^{(\alpha)} = \|(\alpha^n, \cdots, \alpha^1)\|_2(\alpha^{-n}, \cdots, \alpha^{-1})$. The following limits hold for each $z \in \mathbb{C}$:*

$$\frac{\|(zI - A_p)^{-1}\|_d \;\; |p(z)|}{\alpha^{n-1}} = 1 + O\left(\left|\frac{z}{\alpha}\right|^2\right) \quad \textit{as} \quad \alpha \to \infty, \tag{27}$$

$$\frac{\psi(z; d^{(\alpha)})}{|p(z)|} = 1 + O\left(\left|\frac{z}{\alpha}\right|^2\right) \quad \textit{as} \quad \alpha \to \infty. \tag{28}$$

Proof. It can be shown that

$$\phi(z; d^{(\alpha)}) = \frac{\alpha^{n-1}}{|p(z)|}\left(1 + O\left(\left|\frac{z}{\alpha}\right|^2\right)\right) \quad \text{as} \quad \alpha \to \infty.$$

Substituting this expression into Proposition 6.2 establishes (27), and (28) follows from the estimate

$$\|\tilde{z}\|_{1/d^{(\alpha)}} = \alpha^n \left(1 + O\left(\left|\frac{z}{\alpha}\right|^2\right)\right) \quad \text{as} \quad \alpha \to \infty.$$ □

Example 6.4. We may paraphrase Corollary 6.3 as follows: In the limit $\alpha \to \infty$, $L_{\epsilon\alpha^{n-1}}(p)$, $Z_{\epsilon\alpha^{n-1}}(p; d^{(\alpha)})$ and $\Lambda_\epsilon(A_p; d^{(\alpha)})$ agree with one another. Figure 18 illustrates this agreement for the monic polynomial p with zeros $1, 2, 3, 4, 5$. The figure shows $L_{\epsilon\ \alpha^{n-1}}(p)$, $Z_{\epsilon\alpha^{n-1}}(p; d^{(\alpha)})$, and $\Lambda_\epsilon(A_p; d^{(\alpha)})$ for $\epsilon = 10^{-4}, 10^{-3}$, with $\alpha = 20$. The figure shows that the agreement is close, especially between the pseudozero sets and the lemniscatic regions.

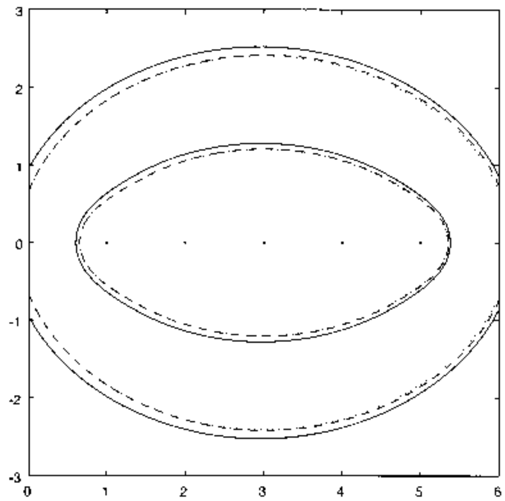


Fig. 18. ϵ-pseudospectra of A_p (solid), ($\epsilon\alpha^{n-1}$)-pseudozero sets of p (dashed) and ($\epsilon\alpha^{n-1}$)-lemniscatic regions of p (dotted) corresponding to perturbations from the scaling operation on p (ϵ = $10^{-4}, 10^{-3}$; α = 20)

7. Conclusions

In this paper we have argued that the pseudozero sets of polynomials tend to match approximately the pseudospectra of the associated balanced companion matrices. This correspondence gives a geometric explanation of why it is that eigenvalue algorithms applied to companion matrices appear to compute polynomial zeros stably, an observation that we have confirmed by experiments comparing the companion matrix code ROOTS (Matlab) with the more traditional zerofinding codes PA16 (Madsen-Reid) and CPOLY (Jenkins-Traub).

Although our results become exact in certain limits (Sect. 6), for most polynomials they are inexact and empirical. We do not claim that they apply to all polynomials without exception. They are solid enough, however, that a reasonable degree of confidence in zerofinding via companion matrices seems justified.

If accuracy is not an argument againist the use of ROOTS, what about speed? We have not discussed the basic fact that whereas matrix eigenvalue algorithms require $O(N^3)$ work, polynomial zerofinding codes such as PA16 and CPOLY come closer to $O(N^2)$. For large N this difference will certainly become significant. However, polynomial zerofinding is a rather unusual problem of scientific computing: one is not interested in large N ! It is very hard to conceive of a genuine application where it would be a good idea to compute zeros of high-degree polynomials specified by their coefficients. Thus it is our view that asymptotic complexity is not of decisive importance for zerofinding algorithms.

References

1. van Dorsselaer, J.L.M., Kraaijevanger J.F.B.M., Spijker, M.N. (1993): Linear stability analysis in the numerical solution of initial value problems. Acta Numerica 199–237
2. Edelman, A., Murakami, H.: Polynomial roots from companion matrix eigenvalues. Math. Comp. (to appear)

3. Gautschi, W. (1984): Questions of numerical condition related to polynomials. In: Golub, G.H., ed., MAA Studies in Numerical Analysis, vol. 24, Math. Assoc. of America, Washington DC
4. Godunov, S.K., Kiriljuk O.P., Kostin, V.I. (1990): Spectral portraits of matrices. Preprint 3, Inst. of Math., Acad. Sci. USSR, Novosibirsk
5. Golub, G.H., Van Loan, C.F. (1989): Matrix Computations, 2nd ed. Johns Hopkins University Press, Baltimore
6. Hille, E. (1987): Analytic Function Theory, vol. 2, 2nd ed. Chelsea Publishing Company, New York
7. Horn, R.A., Johnson, C.R. (1985): Matrix Analysis. Cambridge University Press
8. Jenkins, M.A., Traub, J.F. (1970): A three-stage variable-shift iteration for polynomial zeros and its relation to generalized Rayleigh iteration. Numer. Math. **14**, 252–263
9. Jenkins, M.A., Traub, J.F. (1974): Principles for testing polynomial zerofinding programs. Tech. Report, Department of Computer Science, Carnegie-Mellon University
10. Jenkins, M.A., Traub, J.F. (1972): Algorithm 419 – Zeros of a complex polynomial. Comm. ACM **15**, 97–99
11. Landau, H.J. (1975): On Szegő's eigenvalue distribution theory and non-Hermitian kernels. J. d'Analyse Math. **28**, 335–357
12. McNamee, J.E. (1993): A bibliography on roots of polynomials. J. Comp. Appl. Math. **47**, 391–394
13. Madsen, K., Reid, J. (1975): Fortran subroutines for finding polynomial zeros. Report HL.75/1172(C.13), Computer Science and Systems Divisions, A.E.R.E. Harwell, Oxford
14. The MathWorks, Inc. (1992): MATLAB User's Guide. The MathWorks Inc.
15. Moler, C.B. (1991): ROOTS – Of polynomials, that is. MathWorks Newsletter **5**, 8–9
16. Mosier, R.G. (1986): Root neighborhoods of a polynomial. Math. Comp. **47**, 265–273
17. Ostrowski, A.M. (1959): On the convergence of the Rayleigh quotient iteration for the computation of characteristic roots and vectors, IV. Generalized Rayleigh quotient for nonlinear elementary divisors. Arch. Rat. Mech. Anal. **3**, 341–347
18. Parlett B.N., Reinsch, C. (1969): Balancing a matrix for calculation of eigenvalues and eigenvectors. Numer. Math. **13**, 293–304
19. Peters, G., Wilkinson, J.H. (1971): Practical problems arising in the solution of polynomial equations. J. Inst. Maths. Applics. **8**, 16–35
20. Smith, B.T. (1976): Matrix Eigensystem Routines – EISPACK Guide. Lect. Notes Comput. Sci., vol. 6, 2nd ed., Springer, Berlin Heidelberg New York
21. Trefethen, L.N. (1992): Pseudospectra of matrices. In: Griffiths, D.F., Watson, G.A., eds., Numerical Analysis 1991, Longman
22. Wilkinson, J.H. (1963): Rounding Errors in Algebraic Processes. Prentice-Hall, Englewood Cliffs
23. Wilkinson, J.H. (1965): The Algebraic Eigenvalue Problem. Clarendon Press, Oxford
24. van Dooren, P., Dewilde, P. (1983): The eigenstructure of an arbitrary polynomial matrix: computational aspects. Lin. Alg. Appl. **50**, 545–579

Appendix

The generalized Rayleigh quotient iteration (GRQI) associated with an $N \times N$ matrix A with initial vectors $g^{(0)}$, $h^{(0)}$ and scalar s_0 generates two sequences of vectors $\{g^{(i)}\}_{i\geq 1}$, $\{h^{(i)}\}_{i\geq 1}$ and a sequence of scalars $\{s_i\}_{i\geq 1}$ through the following process:

$$\begin{gathered}
\text{for } i = 1, 2 \dots \\
(A - s_i I)^{\mathrm{T}} g^{(i+1)} = g^{(i)}, \\
(A - s_i I) h^{(i+1)} = h^{(i)}, \\
s_{i+1} = s_i + \tfrac{g^{(i+1)\mathrm{T}} h^{(i)}}{g^{(i+1)\mathrm{T}} h^{(i+1)}}.
\end{gathered}$$

It has been shown that this iteration converges cubically under mild conditions, with the sequence $\{s_i\}_{i\geq 1}$ converging to an eigenvalue λ of A and $\{g^{(i)}\}_{i\geq 1}$ and $\{h^{(i)}\}_{i\geq 1}$ converging respectively to left and right eigenvectors of A associated with λ [17]. For general matrices A, the iteration is carried out numerically by solving two linear systems, for $g^{(i)}$ and $h^{(i)}$, in each step. However if $(A - zI)^{-1}$ is known analytically or certain information concerning the left and right eigenvectors of A is known, this process may be simplified. In this Appendix, we show that if A is the companion matrix A_p associated with a monic polynomial p, then a certain modification of GRQI leads to the Jenkins-Traub variable-shift iteration for polynomial zeros [8]. In other words, the Jenkins-Traub iteration can be interpreted as a scheme for taking advantage of companion matrix structure in a Rayleigh quotient iteration so that the work per step is reduced from $O(N^3)$ to $O(N)$. This observation is originally due to Jenkins and Traub themselves (Sect. 5 of [8]). The following derivation is the reverse of theirs in that the modified GRQI is derived from the variable-shift iteration rather than the other way around.

If λ is an eigenvalue of A_p, then the vector $\tilde{\lambda} = (1, \lambda, \ldots, \lambda^{n-1})^{\mathrm{T}}$ is a left eigenvector of A_p associated with λ. Therefore, instead of having an independent sequence $\{g^{(i)}\}_{i\geq 1}$ for the left eigenvectors, we may modify the GRQI process for A_p by taking $g^{(i)}$ to be the vector

$$g^{(i)} = \tilde{s}_{i-1} = (1, s_{i-1}, \ldots, s_{i-1}^{n-1})^{\mathrm{T}}.$$

The generalized Rayleigh quotient iteration then reduces to the following:

(29) $\quad$ for $i = 1, 2, \ldots$

(30) $$(A - s_i I)h^{(i+1)} = h^{(i)},$$

(31) $$s_{i+1} = s_i + \frac{\tilde{s}_i^{\mathrm{T}} h^{(i)}}{\tilde{s}_i^{\mathrm{T}} h^{(i+1)}}.$$

Further simplification is possible because $(zI - A_p)^{-1}$ is known analytically:

$$(zI - A_p)^{-1} = \frac{1}{p(z)}\tilde{b}(z)\tilde{z}^{\mathrm{T}} - \begin{bmatrix} 0 & 1 & \cdots & z^{n-2} \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix},$$

where $\tilde{b}(z) = (b_0(z), \ldots, b_{n-1}(z))^{\mathrm{T}}$ is the vector of coefficients of the polynomial $(p(w)-p(z))/(w-z) = \sum_{i=0}^{n-1} b_i(z)w^i$. Applying this formula in (30), we get

(32) $$h^{(i+1)} = \frac{H_i(s_i)}{p(s_i)}\tilde{b}(s_i) - \eta^{(i)}(s_i),$$

where $H_i(z)$ is the polynomial $H_i(z) = \sum_{k=0}^{n-1} h_k^{(i)} z^k$ and $\eta^{(i)}(s_i)$ is the vector of coefficients of the polynomial $(H_i(z)-H_i(s_i))/(z-s_i)$. Note that we have used the identity $\tilde{s}_i^{\mathrm{T}} h^{(i)} = \sum_{k=0}^{n-1} h_k^{(i)} s_i^k = H_i(s_i)$ in arriving at the above expression.

Define $H_{i+1}(z)$ to be the polynomial $H_{i+1}(z) = \tilde{z}^{\mathrm{T}} h^{(i+1)}$. Substituting (32) into this expression gives

(33) $$H_{i+1}(z) = \frac{H_i(s_i)}{p(s_i)}\tilde{z}^{\mathrm{T}}\tilde{b}(s_i) - \tilde{z}^{\mathrm{T}}\eta^{(i)}(s_i)$$

(34) $$= \frac{-1}{z - s_i}\left[H_i(z) - \frac{H_i(s_i)}{p(s_i)}p(z)\right].$$

Let $\overline{H_{i+1}}(z)$ be the normalized polynomial of $H_{i+1}(z)$, i.e., $\overline{H_{i+1}}(z) = (p(s_i)/H_i(s_i))\ H_{i+1}(z)$. Since $\tilde{s}_i^{\mathrm{T}} h^{(i)} = H_i(s_i)$ and $\tilde{s}_i^{\mathrm{T}} h^{(i+1)} = H_{i+1}(s_i) = (H_i(s_i)/p(s_i))\ \overline{H_{i+1}}(z)$, (31) is equivalent to

(35) $$s_{i+1} = s_i + \frac{p(s_i)}{\overline{H_{i+1}}(s_i)}.$$

The combination of (34) and (35) is exactly the Jenkins-Traub variable-shift iteration.

Note added in proof. Since submitting this work for publication, we have become aware of some additional references that should be mentioned. The new paper [12] by McNamee consists of a 3500-item bibliography on $3\frac{1}{2}$ inch diskette of papers related to roots of polynomials. W. Kahan has pointed out to us that since

Proposition 3.1 involves matrices, not operators of infinite dimension, its proof can be simplified by the use of dual norms [7] rather than the Hahn-Banach theorem. Finally, our attention has been drawn to a result of van Dooren and Dewilde [24], recently sharpened by Edelman and Murakami [2]. These authors prove that if A is a companion matrix, then any infinitesimal dense matrix perturbation $A + \delta A$ is similar to a perturbed matrix $A + \delta A'$ of companion form, with $\|\delta A'\| \leq C\|\delta A\|$ for some C depending on A but not δA. Edelman and Murakami give an exact formula relating $\delta A'$ to δA based on an elegant argument of transversality. Their theorem provides a precise connection between the zeros of polynomials and the eigenvalues of companion matrices. On the other hand, it is a first-order result, and it does not give explicit bounds on C; indeed, its validity does not depend on whether or not A has been balanced. Also, it does not make a priori predictions about perturbations of polynomial coefficients in the coefficientwise relative sense. For these reasons, it seems there is some work to be done to bridge the gap between the existing theory concerning perturbed companion matrices and the evidence presented in this paper that "ROOTS is stable" in a practical sense.

This article was processed by the author using the LaTeX style file *pljour1* from Springer-Verlag.